

Creating_Good_Documentation

Understanding the global environment depends on the ability of people around the world to share observations and, more importantly, to understand observations made by many partners. This understanding depends critically on the documentation of those observations. If detailed documentation is readily accessible and understandable, the data will be trusted and understood, and integrated into the international data fabric. If that documentation is not available, the data will be collected, but not used.

During the last several years, a series of international (ISO) metadata standards have emerged that will form the foundation for this understanding. The adoption of these standards in the United States will involve a significant transition in the way the U.S. environmental community documents data and in the ways documentation is used by humans and applications. The impacts will extend significantly beyond the data discovery role that has motivated metadata developments in the United States over the last several decades to include detailed documentation of lineage, processing and data quality. The focus will be on ensuring that observations are independently understandable by many diverse users.

The software engineering community has recently been very successful by envisioning and implementing the software development process as a series of spirals, each of which addresses a small set of user requirements. Each spiral involves several phases: requirements collection and prioritization, implementation, testing, and, most importantly, on-going interaction with users. Each spiral builds on previous work and requirements are addressed through a series of on-going iterations, each of which results in a more capable system.

Like a multi-spiral software development process, the creation of high-quality, complete documentation is an on-going interaction between several groups. It is an end-to-end process that encompasses the complete data life cycle. This document identifies those groups and describes their roles in a generalized overview of the documentation creation process.

Contents

- 1 Actors
- 2 Iterations
 - ◆ 2.1 1. Collecting Existing Documentation Resources
 - ◆ 2.2 2. Formulating Standards-Based Documentation Implementation
 - ◆ 2.3 3. Checking Back With Data Collectors/Providers
 - ◆ 2.4 4. Checking Back With Users
 - ◆ 2.5 5. Extending the Breadth
- 3 Potential Spirals

Actors

Successful documentation efforts require input from a variety of groups, each with specific experience and knowledge.

Users are the group that must be able to understand the data they receive and to confidently apply them to problems they are trying to solve. They are the "rubber on the road" and their capability to understand and use the data is the metric used to characterize the success of the documentation process. These people are referred to as the Designated Community in the Open Archival Information System Reference Model (OAIS-RM). This group provides real-world experience using the observations and data and insight into documentation requirements derived from that experience.

Data Collectors/Providers are the group responsible for collecting and processing observations. They design instruments, observing systems and processing systems and operate all three. Some of these people are scientists working on research projects on the cutting edge of environmental science. Some are parts of

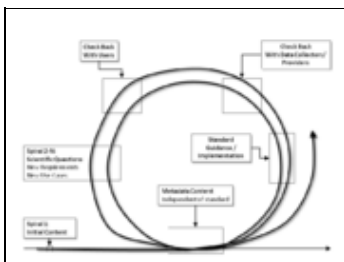
Creating_Good_Documentation

operational teams trying to keep legacy equipment and systems running just a little bit longer! They are the foundation of the documentation process, understanding the sometimes arcane details that affect the observations and creating the original source materials (either physical or mental) for the documentation system.

Data Stewards are the people that take long-term responsibility for sharing observations with users and for ensuring that the users can understand the data they receive. In many cases, they represent the Data Collectors/Providers to the users. They continue in this role after the Data Collector/Provider has moved on to other problems and, sometimes, to other careers or retirement. Preserving data and understanding is a difficult process of communicating with the future. Data Stewards orchestrate that communication.

Heterogeneity is a fundamental barrier to preserving data and understanding across the diverse collection of observations and data required to characterize the global environment. The level, and impact, of heterogeneity can be partially mitigated using standard approaches and practices across a community of users, providers, and stewards. In some cases these approaches are expressed as de jure standards created and maintained by national or international standards organizations. In other cases, the approaches are developed and promulgated as de facto standards or practices by the same communities. Standards Experts are people with experience with de jure and de facto standards and practices that share that experience with the other actors described above in order to facilitate adoption of these standards and reductions in the heterogeneity of the total system. Their contributions cross disciplines as they facilitate horizontal connections across the background of vertical organizations.

Iterations



Documentation Spirals

As described above, the process of creating good documentation is an on-going process made up of a series of spirals or iterations. Each of those spirals includes several phases, described here.

1. Collecting Existing Documentation Resources

The first step in the process of creating high-quality documentation is collecting existing documentation resources and understanding the roles of those resources. These might include Web pages and data access systems, user guides and papers published in journals or on the Web, documentation included in the files used to distribute observations and data, and existing standard documentation. The goal of this initial phase is to collect documentation that is currently discoverable by the group that ultimately depends on this documentation to understand the data, the Users. This task is generally split between the Data Stewards (90%) and the Data Providers (10%).

2. Formulating Standards-Based Documentation Implementation

Once existing documentation is collected and accessible, the Standards Experts come into the picture to characterize the existing documentation and the requirements that it serves. They bring experience with

standard approaches used to address general documentation requirements and apply that experience to the existing documentation. They can identify standard approaches and tools that might be used with the existing documentation as well as gaps in 1) the existing documentation and 2) existing standards and approaches that might be filled in subsequent iterations.

This phase is focused more on the representation of the existing documentation than its content. It is generally split between the Standards Experts (75%) and the Data Stewards (25%). At the end of this phase the Data Stewards should be familiar with the specifics of how the standards have been applied to the existing documentation. They should be able to find the initial documentation content in the standard representation and explain the representation to the Data Collectors/Providers.

3. Checking Back With Data Collectors/Providers

The first iteration results in an initial foundation for a standard representation of the existing documentation and some ideas about how existing documentation and standard approaches might evolve in response to gaps identified in phase 2. Now the Data Collectors/Providers join the team. They serve as fact checkers and elaborators in order to ensure that the initial documentation is correct and takes maximum advantage of existing resources. This task is generally split between the Data Collectors/Providers (60%), the Data Stewards (30%) with the Standards Experts answering questions and providing advice if needed (10%). The focus is on the documentation content rather than on the representation of that content. At the end of this phase, the Data Collectors/Providers should be confident that the content of the existing metadata is included in the new representation.

It is important to remember to include operators of the systems that collect and process the observations and data during this phase. Impacts of the new documentation representation and tools on existing systems and processes must be considered and minimized in order for the new approach to be successful.

4. Checking Back With Users

Once the existing documentation content has been translated into a standard representation and checked for accuracy, the new representation needs to be checked with users to ensure access and usability of the new representation. This phase needs to involve existing users, who need to adapt to the change, as well as new users that find it easier to confidently discover or use the observations and data because of the standardized documentation. This second group is a sample of the future users that are hopefully being served by the standard documentation. They are a group that the Data Collectors/Providers have not previously served, so they may be hard to find. At the same time, they are expected to be the group that initially benefits from this process. They are a good source for stories that demonstrate that the process does actually have benefits! This task generally involves the Data Stewards (70%), the Data Collectors/Providers (20%), and the Standards Experts (10%).

5. Extending the Breadth

At the end of the initial spiral through these four phases the collection of existing documentation has been transformed into a new representation and the content has been checked. Existing and new users have been contacted and polled in order to understand impacts of the transition. Positive impacts have been transitioned into success stories and negative impacts have been characterized and understood. A status quo has been reached that, hopefully, is similar to that which existed prior to the start of the process. Hopefully, new users have been introduced to useful data and documentation and harm to existing systems and users has been minimized.

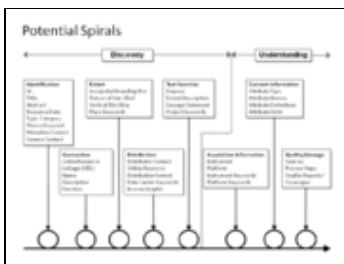
It is almost certain that important questions that are not addressed by the existing documentation will be identified during the four phases described above. These questions make up the pool from which the

Creating_Good_Documentation

requirements for the next iteration are identified. The selection process is informed by input from all of the team members with recognition that the subsequent iteration should target small steps with goals that everyone understands and agrees will provide significant benefits.

The approach is similar to that outlined above with a couple of exceptions. The first spiral started with existing documentation and the roles/requirements that it served and the assumption that those requirements were addressed by the existing documentation. Subsequent spirals start with a small set of priority requirements that are not addressed by the existing documentation. Content required to address those requirements is identified by the team and the Data Collectors/Providers are enlisted to identify or develop sources for that content. The Standards Experts examine the content and suggest either approaches to a standard representation or extensions to existing standards or practices that might accommodate that content. This is equivalent to Phase 2 described above and the rest of the process is similar to that in the first spiral.

Potential Spirals



Potential Spirals

The content and order of the documentation spirals is related to the specific scientific needs and requirements of particular groups. A helpful first step for any group is to identify content that exists in all records (see [ISO Boilerplate](#) for some suggestions). See [Documentation Spirals](#) for other possibilities shown in this Figure.